AD_____

Grant Number DAMD17-96-1-6182

TITLE: Genes Altered by Intracisternal A Particles in Mouse
Mammary Tumorigenesis

PRINCIPAL INVESTIGATOR: Michael R. Crowley

CONTRACTING ORGANIZATION: Health Research, Incorporated
                            Buffalo, New York   14263

REPORT DATE: July 1997

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                Fort Detrick, Maryland   21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
                          distribution unlimited

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.

19971210 034

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1997	Annual (1 Jul 96 - 30 Jun 97)

4. TITLE AND SUBTITLE
Genes Altered by Intracisternal A Particles in Mouse
Mammary Tumorigenesis

5. FUNDING NUMBERS
DAMD17-96-1-6182

6. AUTHOR(S)
Michael R. Crowley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Health Research, Incorporated
Buffalo, New York 14263

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200

The purpose of the original proposal was to search for and isolate new transcripts associated with mammary carcinogenesis. In order to accomplish this we have been analyzing mammary carcinomas induced by 7,12-dimethylbenz (a) anthracene (DMBA) in BALB/c mice which express high levels of intracisternal A-particles (IAP). Differential hybpridization and differential display strategies are being used to isolate transcripts which contained IAP LTR sequences. A clone was isolated by library screens contained two poly adenylation signals and poly A tracts suggesting a rearrangement in library construction. Serendipitously, the unique portion of the clone hybridized to two transcripts upon Northern analysis, one tumor specifically. All tissues examined express the larger transcript while tumors of various etiologies express the smaller tumor associated transcript. Backcross analysis has localized the transcript to mouse chromosome 17.

14. SUBJECT TERMS Breast Cancer
Intracisternal A Particle, novel cancer gene, mammary
chemical carcinogenesis, DMBA

15. NUMBER OF PAGES
34

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_ue_ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_ue_ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_ue_ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

_ue_ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_ue_ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature   7/31/97
                                                  Date

TABLE OF CONTENTS

INTRODUCTION

IAPs are members of a multigene family found in approximately 1000 copies per haploid geneome of the mouse *Mus musculus* as well as most other rodents (1). They are defective retroviruses which contain 3' and 5' long terminal repeat (LTR) sequences and coding regions for gag and pol genes. IAPs do not however, encode a functional envelope (env) protein due to the presence of several inframe stop codons. IAPs are produced in the cisternae of the endoplasmic reticulum (ER) as virion-like particles consisting of an outer shell composed of the ER membrane with an electron dense layer surrounding a translucent center. These particles have not been found extracellularly, and horizontal transmission of IAPs has not been demonstrated.

Transcriptional activation of the IAP molecule is critical for it's ability to cause mutation via retrotransposition. Recently it has been discovered that activation of c-Ha-*ras* promotes "global" hypomethylation in murine embryonic P19 cells (2).This finding is intriguing in that epigenetic factors, such as methylation status, can cause an alteration in IAP expression (1, 3). Specifically, demethylation of the IAP promoter is associated with increased IAP expression. We have found that *ras* is indeed mutated at codon 61 in 100% of the primary DMBA mouse mammary tumors analyzed to date and that IAP expression is elevated in all of the same tumors (our unpublished observations).

Retroviruses, both endogenous and exogenous, can be important sources of genetic variation within an organism. IAPs have been shown to retrotranspose and insert into new locations in the genome at $\sim1.5\times10^{-6}$ per cell per generation (4). Reintegration of IAPs back into the genome may alter the regulation of nearby cellular genes by providing a promoter or an enhancer to the endogenous gene, or by disrupting regulatory sequences and interrupting normal cellular control over its promoter. Integration may also disrupt the resulting protein by inserting sequences within the RNA which may ultimately affect protein function. Donation of sequences within the transcript may cause selection of the polyadenylation signal of the IAP LTR resulting in a truncated protein with altered function. Indeed, many IAP molecules have been discovered affecting nearby cellular genes in neoplasias (5-11). IAPs also contribute to the normal regulation of cellular genes through the promoter sequences in their LTRs (12). Asch and Asch found that high expression of IAP RNA and protein is present in many mouse mammary tumors and preneoplasias, whereas little or no expression is detected in normal mammary glands from virgin, pregnant, lactating or involuting mice (13,13,15). Therefore changes in IAP expression frequently occur during the progression to tumoriagenesis but not during normal growth and differentiation cycles of the mammary gland (14,15).

In order to isolate novel genes associated with mammary carcinogenesis the method of Mager and colleagues was used. Briefly, overexpression of IAP elements leads to retrotransposition and insertional mutagenesis. IAP elements may contribute sequences to the flanking cellular gene and it is possible to isolate those chimeric transcripts using IAP sequences. In the original proposal two different techniques were described to isolate the chimeras, differential hybridization of a cDNA library made from a DMBA-induced primary mouse mammary carcinoma and differential display from the same type of tumor. Isolation of chimeras, sequence analysis and characterization of the fusion transcripts was to be carried out in the first year of the proposal. The normal homologue of the unique portion of the chimera was proposed to be isolated and characterized. Isolation of the normal counterpart was to be done by library screening and PCR techniques. Southern analysis and isolation of the genomic locus will be used to determine IAPs role in mutagenesis. Finally, mouse mammary tumors from other etiologies will also be examined for the presence of the identified cellular gene.

Previously, we reported that IAP transcripts are increased in several different mouse neoplasias (15). IAP expression must precede reverse transcrpition and reintegration of the element into the genome and the resulting integration may affect the expression and/or structure of neighboring genes. For this reason I have been utilizing two approaches to isolate novel genes from DMBA-induced mouse mammary carcinomas which express high levels of IAPs (data not shown). A cDNA library was constructed from the DMBA-induced mammary carcinoma which expressed the highest levels of IAPs and screened this via a differential hybridization method (Fig. 1). Dixie Mager and her colleagues have isolated chimeric RNA molecules using the same technique from a

NTeraD2 human cDNA library with probes specific to Human endogenous retrovirus, HERV-K (19). In the prliminary results to the original proposal, the library was screened with three different probes corresponding to different portions of the IAP element (Fig. 1) The first hybridization was with the envelope gene to identify all IAP elements within the library. The second hybridization was with a specific probe to the U3 region of the IAP LTR. This probe would identify the same transcripts as the env probe, it would also identify solo LTRs expressed and potentially any transcripts which contained an LTR but not the rest of the IAP element. The third probe was specific to the U5 region of the LTR and would identify similar transcripts as the U3 probe. Transcripts containing U3 sequences could represent chimeric molecules in which the U3 either donated the polyadenylation signal or were spliced into the body of the RNA transcript (Fig. 1B and C). The U5 probe could identify those chimeras that were promoted by the LTR or spliced into the body of the transcript (Fig. 1B, D). Efforts have focused on those chimeras that were positive for the U3 region and negative for the *env* region (Fig. 1C). At the time of writing the original proposal, 30 clones which were positive with the U3 region, two of which contained the appropriate structure. Five clones did not continue to hybridize with the LTR U3 probe after plasmid rescue, several inserts were 300bp or less and when sequenced were determined to be LTR's only, and three others were the same clone which contained only IAP gene sequences.

EXPERIMENTAL METHODS:

Tumors and RNA analysis. 7,12-dimethylbenz (a) anthracene was given to virgin BALB/c mice beginning at 8 weeks of age at 1mg/dose/week for 4 weeks. The animals were kept as virgins to induce ductal hyperplasias which eventually progressed to ductal carcinomas. The animals were sacrificed by CO_2 asphyxiation and the tumors removed and stored in liquid nitrogen. A piece of each tumor was removed and RNA was isolated via the acid-phenol guanididium isothiocynate method (Chomchinski and Sacchi). The various mouse tissue RNAs and the TM series of cell lines analyzed were extracted using TRI reagent (Molecular Research Center, Inc. Cincinnati, OH) on fresh tissue or placed directly on cells. Ten micrograms of RNA was electrophoresed through a 2.2M formaldehyde gel and transferred to nylon membrane (Zeta Bind). The RNA was fixed to the membrane by UV cross linking at 1600x100μJ constant energy. Hybridizations were done in Church and Gilbert hybridization solution (0.5M $NaPO_4$ pH 7.4, 7% SDS, 1%BSA and 0.001M EDTA) at 65°C then washed in 2XSSC for 10 min at room temperature, and stringently in 0.2XSSC, 0.1%SDS at 58°C for 30 minutes with one change for a total of 1 hour. The blots were exposed to XOMAT AR-5 X-ray film (Kodak) at -70°C with intensifying screens for 3 days to 2 weeks. Blots were stripped in 2mM Tris [8.0], 1mM EDTA and 0.1%SDS at 70°C for 30 min. All hybridization probes were random primed DNA probes (Feinberg and Vogelstein). Polyadenylated RNA for library construction was isolated by oligo dT column chromatography (Molecular Research Center, Inc. Cincinnati, OH).

cDNA library screening and Differential Display analysis. A cDNA library was constructed as described in the original proposal. Briefly, poly A^+ RNA was isolated from the tumor which expressed the highest levels of IAPs (tumor 12.21). First and second strand cDNA synthesis and ligation into the λ ZAP cloning vector were done as described by the manufacturer (Stratagene). The library was amplified once and then screened by filter hybridization with ~50,000pfu/filter. Differential hybridization was done with the following probes: a PCR generated IAP U3 specific amplicon and a restriction fragment from MIA14.1 representing the env gene of IAP. Films were compared and those areas which hybridized to both probes were eliminated as they represented IAP transcripts. Those areas which hybridized to only the U3 specific probe were analyzed further. A PCR generated probe specific for the U5 region of IAP was also used as a probe but those clones have not been purified further. In an attempt to isolate the full length normal transcript a λZAP mouse kidney cDNA library and a λgt10 mouse liver cDNA library (generous gifts from Dr. Deborah Nagle, Millenneium Pharmaceuticals, Cambridge, MA) were screened as describe above.

Differential display was done essentially as described by Liang and Pardee with some modifications (16). Briefly, 0.2μg of DNA-free total RNA isolated from mammary glands from virgin, pregnant and lactating mice as well as from three DMBA induced tumors (T6.8, T719A, T12.2A from our laboratory or T4657 from D. Medina, Baylor College of Medicine, see Figure X) were ethanol precipitated with the downstream primer IAP U3Seq2 5'GCGCAGATTATTTGTTTACC3'. The RNAs/primers were resuspended in DEPC-H_2O and 1X RT buffer and incubated at 65°C for 5 minutes to remove any secondary structure then transfered immediately to 55°C for 10 minutes to anneal the primer. Reverse transcription was carried out as essentially described by the manufacturer of the RT (Gibco/ BRL) in a 20μl reaction volume for 45 minutes at 50°C (to increase specificity of the reaction). The reactions were heated to 65°C for 15 minutes to kill the RT. Two microliters of each RT reaction were used in the subsequent PCR reaction.

Upstream primers were either obtained from GeneHunter (Mass) or modified from the original AP primers (GeneHunter) by the addition of one nucleotide on either end of the primers. The AP-M primers were 12mers (Curachem) of the following sequences; Ap-1m: 5'CAGCCAGCGAAG3', AP-2m 5'AGACCGCTTGTA3', AP-3m 5'TAGGTGACCGTC3', AP-4m 5'AGGTACTCCACG3', AP-5m 5'TGTTGCGATCCA3' . The PCR reaction conditions were 10mM Tris (8.4), 50mM KCL, 1.5mM MgCl$_2$, 80μM dNTPs, 2μCi ^{32}P-dCTP, and 0.5units of Taq DNA polymerase (Boehringer Mannheim, Indianapolis, IN) or 10mM Tris (8.4), 50mM KCL, 0.1% Triton-X-100, 1.5mM MgCl$_2$, 80μM dNTPs, 2μCi ^{32}P-dCTP and 0.5 units of Taq DNA polymerase (Promega, Madison, WI)with cycling parameters of 94°C for 45 sec/ 45°C for 2min/ 72°C for 30 sec for 40 cycles. The PCR reactions were run on a 6% polyacrylamide denaturing gel at 50W for 3.5 hours. The gels were dried and exposed to XOMAT AR-5 X-ray film at -70°C for 12 hours with intensifying screens. Those bands which were unique to a tumor or to the series of tumors or were up or down regulated compared to the three normal tissues were excised from the gel and rehydrated in water overnight. Reamplification of the bands was done by simply removing the chromatography paper attached to the gel slice with forceps and cutting a piece of the gel slice. The gel slice was added to the PCR reaction with the primers used to amplify the original band, using the same reaction conditions as described for differential display. The amplified products were run on a 2% agarose gel and gel purified using either the MERmaid isolation kit or the Geneclean isolation kit (both Bio 101) depending on the size of the amplicon. The purified DNA was used as a template for random priming and hybridization to Northern strip blots containing a normal tissue RNA and a tumor RNA.

cDNA cloning. All clones generated from the tumor specific λZAP cDNA library or the mouse λZAP kidney cDNA library were excised by a helper phage method described by the manufacturer. Lambda clones isolated from the mouse liver specific library were plaque purified and the DNA isolated via liquid lysate (17). The insert was liberated from the λgt10 vector by EcoRI digestion. The insert was purified from a 1% agarose gel using Geneclean (Bio 101) and ligated into the EcoRI site of pGEM7Z+. Differential display products were cloned into the pGEM Easy T/A cloning vector (Promega).

Sequence analysis. All sequencing was done by the Sanger dideoxy chain termination method. Much of the sequence analysis was performed in this laboratory by 33-P autoradiography while limited sequence analysis was done in Roswell Park's core facility using the ABI automated sequencer. Sequencing primers were the M13 universal forward and reverse primers as well as a sequencing primer from the U3 region of IAP (5'GAACACAGCTGTCAGCGCC3'). All sequences were analyzed using the GCG program maintained by the University of Wisconsin at Madison.

Southern analysis. Isolated cDNA clones from the λZAP tumor library were digested with EcoRI and XhoI to excise the inserts. The DNAs were run on a 1.5% agarose gel, denatured, neutralized and transferred to nylon membranes by standard techniques. The Southerns were hybridized in Church and Gilbert hyb buffer (see above) using the same probe as were used for the library screen (see above). Those clones that continued to hybridize to only the U3 specific probe were sequenced.

Genome Localization of Kokopelli. A Mus musculus (C57BL/6J) X Mus spretus backcross was obtained for The Jackson Laboratory (Bar Harbor, Maine) and used for localization of the p17b(kokopelli) transcript. Briefly, 5 ng of genomic DNA was subjected to PCR using the primers p17bGSP2F 5'CCCAGTATGACTCCCAGCAGGC3'

and p17bGSP3 5'AAAGTGGACCTGAATGACACAC3' in a "hot start" reaction which consisted of 10mM Tris (8.4), 50mM KCL, 0.1% Trition-X-100, 1.5mM $MgCl_2$, 200μM dNTPs, and 0.1 unit of Taq DNA polymerase (Promega) in a 20μl reaction volume overlaid with mineral oil. The PCR cycling parameters were 95°C, 5 min/ 80°C hold (addition of all reagents to the denatured DNA occurred within the hold setting) then 35 cycles of 94°C for 45s/ 57°C for 45s/ 72°C for 30s. These primers amplify a 410bp product which was found to contain a Hae III polymorphism between C57BL/6J and spretus. The amplicons were digested with Hae III and run on a 2% agarose gel and stained with ethidium bromide. Hae III digests the C57BL/6J allele to yield an ~215bp fragment, an ~110bp fragment and an ~ 85bp fragment. The size of the spretus allele was not altered by digestion with Hae III. A total of 96 meiosis were screened in this PCR based assay but only 91 animals gave a scorable result. The 5 remaining DNA samples were degraded and gave no PCR products. Therefore recombination frequency was based on the number of recombinants divided by the total number of DNA samples which gave PCR products. Chromosome assignment and recombination frequency (or map distance) was determined using the program Map Manager v2.6.5 (K. Manly, Roswell Park Cancer Institute).

<u>5' Rapid amplification of cDNA ends (RACE)</u>. Total RNA was subjected to reverse transcription using the gene specific primer p17bGSP1 5'GTCACAAAGGAACAGGC3'. Reverse transcription was done with Superscript RT (Gibco, BRL) as per the manufacturers directions. The first strand was digested with RNase H and an oligo dC tail was added to the 5' end via terminal deoxytransferase according to the manufacturer (Gibco, BRL). One fourth of the tailing reaction was used in a PCR reaction with the 5'RACE anchor primer 5'CUACUACUACUAGGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG3' from BRL and the primer used to generate the first strand cDNA. A nested PCR reaction followed the initial PCR reaction using the universal amplification primer 5'CUACUACUACUAGGCCACGCGTCGACTAAGTAC3' and p17bGSP2 5'ATGGAAAACGGGATATAACGAA3' in 10mM Tris (8.4), 50mM KCL, 1.5mM $MgCl_2$, 200mM dNTPs, 0.5μM each primer and 1 unit of Taq polymerase in 50μl.

5' RACE was also done according to Troutt (18). Briefly, first strand cDNA was synthesized as previously described (see above). The RNA was removed by using both RNAse H and RNAse A and the single strand cDNA was purified using NACS column chromatography as described by the supplier (Gibco, BRL). The cDNA was ethanol precipitated, then resuspended in buffer (50 mM Tris (8.0), 10 mM $MgCl_2$, 10mg BSA per ml, 25% (wt/vol) PEG 8000, 1mM hexamine colbolt chloride , and 20μM ATP, in 10μl volumes) and 10pmol of an oligonucleotide primer 5'CTCTCCCTTCTCGAATCGTAACCGTTCGTACGAGAATCGCTGTCCTCTCCTTY GAGCT3'. The primer was blocked on the 3'end by TdT transfer of ddATP and phosporylated on the 5' end with T4 polynucleotide kinase (NEB). T4 RNA ligase (New England Biolabs) was added and the reactions incubated at 22°C for 12-18 hours. The reactions were then subjected to two rounds of nested PCR with primers LAP-2 5'AGCGATTCTCGTACGAACGG3'/p17bGSP-1 (see above for sequence) and LAP-1 5'ACGAACGGTTACGATTCGAG3'/p17bGSP-2 (see above for sequence) respectively. Other sets of primers were also used to isolate the 5'end of the clone i.e. LAP-2/p17bU3S6 (5'ACACCAGCTGTGCTTG3') and LAP-1/p17bGSP3R (5'GACCAGGCTGGCCTCAACAC3').

A third method of 5'RACE was also tried. Poly A+ RNA form tumor 6.8 was converted into cDNA using oligo dT as the first primer, Eco RI linkers were ligated onto the ends with T4 DNA ligase (BRL), the linkers were digested with Eco RI and the cDNA purified by Sephadex G50 column chromatography. The 6.8 cDNA was ligated into pGEM7Z+ previously digested with

Eco RI. The ligation mixture was subjected to PCR with primers p17bGSP2F and p17bGSP3 to determine if the target molecules were present. 5'RACE was accomplished by using a vector specific primers, -47 Forward primer or the M13 reverse primer and gene specific primers p17bGSP5 (5'AGGACGGTGACAACACACAT3') and p17bGSP3. Two rounds of nested PCR were completed with the above primers using the reaction conditions described above for hot start PCR with cycle parameters as follows: 35 cycles of 94°C for 45 sec, 58°C for 45 sec and 72°C for 30sec. The PCR products were run on a 2% agarose gel, excised, geneIceaned and cloned into pGEM EASY T/A (Promega) as described above. Another set of nested gene specific primers (p17bU3S6 and p17bGSP3R) were also used with similar parameters except the annealing temperature for p17bU3S6 was 45°C and that for p17bGSP3R was 58°C.

RESULTS AND DISCUSSION

Isolation of Novel Chimeric Molecules via Differential Hybridization.

As stated in the preliminary results of the proposal and in the Introduction, the two other original clones had LTR sequences but also contained a limited amount of the *env* gene attached to the unique portion of the clone. This might suggest that these chimeras arose from mechanisms other than retrotransposition. The unique portions of the clones, p13a and p17b, were used as probes on Northern blots. p13a was expressed in all tissues examined (data not shown). Furthermore, its transcript size was not altered in any tumor expressing IAPs nor was there any change in expression levels (Fig 2). This clone was not further analyzed. The other clone isolated was p17b. Restriction digest allowed removal of the IAP sequences to make a probe from the unique portion of the clone (Fig. 3A). Northern analysis showed the unique portion hybridized to two transcripts, an ~6.6knt found in all tissues and an ~1.4knt found only in the tumors (Figure 3B). The preliminary results suggested that the technique was working, in that cellular sequences chimeric with IAP sequences were obtained. From the initial studies, p17b appeared to be a good candidate gene to isolate due to it's expression pattern in DMBA-induced mouse mammary carcinomas (Fig. 3B).

To obtain full length cDNA clones of p17b the λ12.21 library was rescreened with the 1.3kb fragment of p17b. Several clones were isolated and sequenced and one was found to have the same 3' end as p17b. This clone p17b-20 was sequenced fully and was determined to be the full length clone of U1 snRNP specific protein C (Fig. 4A) This clone was also used as a probe on Northern blots and it was determined that it was not overexpressed or misexpressed in RNA from tumors compared to normal tissues (Fig. 4B). Concurrently, I replated the tumor library and screened it again with IAP LTR probes to isolate any clones that may have been missed in the initial screen. Several more clones were isolated and were found to be chimeric with IAP. However, the clones were either not informative (i.e. expression patterns between normal and tumor was not altered) or were deemed to be rearrangements within the library (data not shown and Fig 5). For example, clone p48 was found to contain chimeric sequences with another gene. p48 had a similar structure as p17b due to the presence of two stretches of poly A within the same clone. The downstream poly A tract was not preceded by the usual AAUAAA, however. Indeed, sequence analysis reveals that the homology is within the gene and not at it's 3' terminus. Northern analysis of p48 revealed that in 30% of tumors analyzed the transcript did indeed have an altered pattern of expression (data not shown and Fig 5B). When sequenced, the clone was found to be 100% homologous to vascular smooth muscle alpha actin (α-actin) while the sequence between the poly A tracts was homologous to nucleotides 1672 to 1909 of a murine hydrophobic mitochondrial protein (Fig. 5C and data not shown). This result was not explored further. It did however put into question the quality of the library and if there had not been severe rearrangements rendering the library unusable.

Sequence Analysis of Isolated Clones.

Sequence analysis of p17b revealed that the IAP sequences associated with the unique region probably arose due to recombination within the library. Figure 6 shows that the 3' end of the clone is indeed IAP LTR and has the conformation that one would predict for 3'end truncation (i.e. donation of poly A signal and tail). However, the clone also contained another poly A signal and a second poly A tail as well as sequence which was 100% homologous with U1 snRNP specific protein C. From the Pst I site (Figure 7) the sequence was unrelated to anything in the data bases. There was homology to the murine B1 repetitive element from but no homology to anything in the database from the begining of the clone to the B1 element. The Northern analysis described above was done using a 1.3kb Eco RI fragment from p17b. This Eco RI fragment represented all the sequence elements found in the clone, therefore I designed PCR primers to isolate regions within the larger clone. Restriction digests of the clone was used also to isolate further subclones to use as probes. PCR with primers specific to the junction between the U1 snRNP and the upstream unique region (p17bGSP6 and p17bGSP4, respectively) (see methods and materials and Figure 6B) were used to generate a probe template. This probe gave exactly the same hybridization pattern as that

7

observed with the larger 1.3kb probe. Other PCR primers were used in similar experiments corresponding to the B1 element and the unique downstream region (p17bGSP2F and p17bGSP3, respectively) with identical results. The unique region of the clone, between the B1 element and U1 snRNP, isolated from the 1.3kb probe by Bgl II/Pst I double digest gave the same results as the three previous probes (data not shown and Fig 7A). These data would suggest that the clone, up to but not including the IAP molecule, has a similar structure in the tumor and normal cells as it does in the library. Attempts to generate PCR amplicons via RT-PCR analysis across the snRNP/IAP junction have not produced bands sizes which would be predicted by the cDNA clone. Therefore I suggest that the original association with the IAP molecule may have arisen due to recombination within the library during library amplification.

5'and 3' RACE to Identify Full Length Tumor Associated Transcript.

In addition to rescreening the tumor specific library, I attempted 5' rapid amplification of cDNA ends (RACE) with p17b primers using tumor and normal RNA. The BRL 5' RACE kit was used with gene specific primers p17bGSP1 and p17bGSP2 (sequences are in Methods and Materials). The specific primer p17bGSP1 was used in the initial reverse transcription reaction. This RACE method is based on terminal transferase addition of dCTP to the 3' end of the first strand RT product and two rounds of PCR amplification, one nested (p17bGSP2 as the primer). I was unable to obtain PCR amplicons under any conditions. The RACE technique of Troutt et. al. (1992) was used. This method utilizes T4 RNA ligase to add a single stranded oligonucleotide linker to the first strand cDNA product. PCR amplification with primers specific to the linker and the gene are used with a nested PCR reaction to increase specificity. This method also gave no results under various reaction conditions. A variation of 5' RACE was then done; cDNA prepared from tumor 6.8 was ligated into the Eco RI sites of pGEM7Z+, in effect making a plasmid based library. PCR primers specific to the gene (p17bGSP3, p17bU3S6 and p17bGSP3R) were used in conjunction with primers specific to the vector (either the -47 M13 forward primer or the universal M13 reverse primer). Nested PCR was done to ensure specificity of the reaction. Amplicons derived by this method were cloned into the T/A Easy cloning vector (Promega), and sequenced. Sequence analysis of several clones representing both the 3' end and the 5' end revealed 100% homology to the p17b clone isolated from the library. In other words, I isolated clones "directly" from the RNA, without propagation in any bacterial cell, and obtained clones with similar structure as that isolated from the library which I suggested to have recombined. However, all the 5' end clones isolated with method all stooped within the B1 element around position 270 (numbering is from the library isolated clone) (Fig. 7A and B). I have not been able to extend the sequence beyond that obtained to date.

Due to the possible rearrangement of the cDNA library the actual 5' end of the clone is questionable. Therefore primer design has been a problem in that RACE protocols have, in general, a size limitation. RACE can easily amplify products of a few hundred base pairs but it is more difficult to amplify several thousand base pairs. 5' RACE of p17b (kokopelli, see below) is further complicated due to the high level of homology between B1 elements. RACE experiments which employ primers within the B1 element have the potential to amplify other transcripts which contain B1 elements (for example of B1 containing genes see 35). Also, the actual location of the murine B1 element or the unique region between the B1 element and the snRNP are unknown within either the full length transcript or the tumor specific transcript. Subsequently I may have trouble in obtaining RACE products due to large distances. However, the data would suggest that the probes I am using for Northern analysis reside in the 3'end of the gene since they hybridize to both the large and the small transcripts. That is, if I assume that the smaller transcript results from either an internal deletion of the large transcript, or alternate splicing has occurred or if an alternate promoter is used within the gene. As in all PCR experiments, specificity of the reaction is dependent on annealing temperatures of the primers as well as magnesium concentration. By using such disparate primers, i.e. oligo dT for 3' RACE or homopolymer tailing in 5' RACE and a gene specific primer, specificity is compromised due to the low annealing temperature of oligo dT primer (typically 42-45°C). Therefore several rounds of nested PCR may be needed circumvent the problem. Again, primer selection is more difficult and several different primer combinations may be necessary to obtain the correct clone.

8

Northern Analysis of p17b.

Since the large transcript of p17b was expressed in normal mammary glands, brain, liver and kidney I wanted to determine if other tissues expressed p17b. I isolated RNA from various mouse tissues including but not restricted to adipose, brain, whole blood, lung, liver, spleen, stomach etc. The large, 6.6knt, transcript was found expressed in every tissue analyzed albeit to varying levels. The smaller 1.4knt transcript was not found expressed in normal tissues (Fig 8A). Next I wanted to determine if the smaller 1.4knt transcript was unique to DMBA-induced tumors or if it was found associated with tumors of other etiologies. Surprisingly, I observed hybridization to the 1.4knt transcript in all mouse tumors as well as in a non-neoplastic mouse cell line, NOG-8. I did not obtain hybridization to any human tumor cell line tested (Figure 8B). A series of mouse hyperplastic and neoplastic cell lines, the TM series of cell lines a gift from Dr. Daniel Medina, also demonstrated the lower, tumor associated transcript. The 1.4knt transcript was not detected in normal mammary epithelial cell lines EL11 and EL12 by Northern analysis (data not shown). Therefore, I suggest that the alteration in the large 6.6knt transcript or specific expression of the smaller 1.4knt transcript may be a very early event in cellular transformation from a normal state to a pre-neoplastic state.

Genetic Localization of p17b (Kokopelli).

In order to better understand the nature of the potential mutation found in the tumors I decided to localize p17b within the mouse genome. I screened a Mus musculus musculus X Mus spretus ((C57BL/6JxSpetus) X Spretus) backcross from the Jackson Laboratory via PCR analysis with primers p17bGSP2F and p17bGSP3 as described in Methods and Materials. The data was analyzed using Map Manager developed by Ken Manly (Roswell Park Cancer Institute). I observed 3 recombinants between p17b 2F/3 and an anonymous expressed marker D17Wsu92e in 91 animals analyzed. This linkage had a lod score of 18.1. I have tentatively renamed p17b to Kokopelli. These data suggest that Kokopelli is located on mouse chromosome 17 approximately 12.5 cM distal to the centromere (Fig. 9). There are no other proto-oncogenes or tumor suppressor genes located in this region.

Differential Display to Isolate Novel Chimeric Transcripts.

Since I felt that the library had recombined, or at the very least that the IAP molecules within the library were highly recombinogenic, I searched for chimeric molecules via differential display (DD). I modified the original protocol by incorporating a specific IAP primer as the initial primer used in reverse transcription instead of the anchored oligo-dT primer. A further modification was the addition of 2 additional nucleotides to the upstream arbitrary primers to allow an increase in annealing temperatures and specificity of the reaction. Since the mammary gland is in different developmental states and expresses different transcripts normally, I chose to include the three stages which represent resting, proliferating and differentiated epithelium (virgin, pregnant and lactating respectively). IAPs are normally expressed at low levels in mammary glands from pregnant mice (28-30 and our unpublished observations). However, IAPs are either not expressed in mammary glands from virgin and lactating mice or are expressed at undetectable levels by Northern analysis (28-30). This representation will help identify transcripts which are expressed in the tumors and not the normal tissues despite differences in physiological states of the mammary gland. Since a specific IAP primer was used in the RT reaction only IAP related molecules should be detected by this method. The same downstream primer was used in all DD experiments. A further control was to include ~500bp of the 3' end of IAP which includes the U3 region and part of the *env* gene to eliminate IAP transcripts. Differences were not only detected between the normal tissues and the tumors but also within the normal samples and within the tumors samples, as would be predicted due to differences in physiology (Fig 10 A and B). The size range of the amplified PCR products were from >500bp down to <100bp, thus giving a good representation of the RNA transcripts in the different samples. An approximately 500bp differential display product unique to the tumors was obtained using primers AP-3 and IAPU3Seq2 (Arrowhead in Fig. 10A). The purified fragment was radioactively labeled and used as a probe on a Northern strip blot. All other amplicons isolated were treated in the same manner. The AP-3 500bp product hybridized to two bands, 7.2knt and 5.4knt,

only in the tumor (Fig. 11). This DD product displays the exact pattern of hybridization as MIA14.1 (the full length IAP probe). Therefore I concluded that this amplicon was IAP. This result suggested that the modification of using the IAP specific primer allowed amplification of IAP related transcripts. Other DD products also gave similar hybridization patterns and were deemed IAP molecules and not further analyzed (Fig 11). Unfortunately, most other transcripts analyzed gave no detectable hybridization above background or expression was not restricted to the tumor. Figure 11 is representative of Northern strip blots hybridized with various IAP-DD amplicons.

      Differential display analysis has a propensity to give false positive and false negative results. Therefore it has been necessary to analyze many differential display amplicons before finding one or two which are real. The ability to use specific primers for reverse transcription could possibly reduce the amount of false priming and worthless results. Since I was looking for chimeric transcripts to which IAP had donated sequences to a cellular transcript (see Introduction), a primer within the U3 region should have yielded more specific results. As described above I have indeed been able to clone IAP transcripts using differential display. However, to date, I have been unable to isolate novel, IAP containing transcripts. Several factors could contribute to this. First, it is possible that the IAP/cellular transcript chimeras do not exist. This is unlikely due to the high level of IAP expression in the tumors and because IAPs have been documented to retrotranspose and contribute sequences to cellular genes. Second, the primer annealing reaction to total RNA may not be specific enough. The annealing conditions initially are very specific in that annealing is done just below the Tm of the primer. However mis-priming may still occur at that temperature. In addition, the RT primer is not removed before the RT reaction and may mis-prime at the optimal temperature for the RTase. Third, removal of erroneous results can be accomplished by Northern analysis of the isolated amplicon. Problems in radioactive labeling and hybridization of small molecules may contribute to ambiguous Northern results and elimination of possible chimeras. Finally, the arbitrary upstream primers can act as both the upstream and the downstream primer thus adding to the pattern obtained in differential display but having nothing to do with IAPs. Differentiation of these molecules can only be accomplished via cloning and sequencing the products.

      Increasing specificity of the DD reaction is the limiting factor in obtaining reliable results. Removal of the excess RT primer as well as increasing the annealing and RT reaction temperatures may increase specificity. I have successfully done the RT reaction at 55°C. This gives much less product but it's more specific (our unpublished results). Also, including a second IAP specific primer internal to the RT primer in the PCR step may decrease the amount of misleading results.

10

RECOMMENDATIONS TO THE S.O.W.

During the first 12 months of this project DMBA-induced mammary carcinomas were to be analyzed by differential hybridization of a cDNA library and by differential display. Also, within the first year I proposed to sequence the isolated clones, determine their structure and to isolate full length tumor associated transcripts. The later aims were predicated on the fact that obtaining clones would be successful. As described in the results cursory examination (Southern analysis with IAP probes) of isolated clones demonstrated chimerism with IAP. More detailed analysis of the clones (sequence and Northern analysis) revealed the clones were possibly a result of cloning artifacts and unique regions were not tumor specific. However, I was successful in isolating a clone which was tumor specific and found expressed in many tumors arising from different etiologies (an aim for months 30-36). Sequence analysis of this clone was initially done by myself in our lab and subsequent sequence was obtained from Roswell Park's Biopolymer Core facility. This then addresses two of the aims in Task 1 for the first 12 months of the program.

The third aim of Task 1 was the isolation of the full length tumor associated transcript by rescreening the library and/or by RACE techniques. I did rescreen the tumor specific library with the p17b probe but the clones I isolated did not hybridize to the kokopelli transcripts on Northern analysis. These results, along with the potential problems with the library (recombination), suggested isolation of kokopelli using PCR techniques would be a better strategy. With the sequence information available at the time, 5' to the B1 element, PCR primers were designed to obtain the 5' end of the clone. RACE procedures using a plasmid based cDNA library with primers to the vector and primers either within the B1 element or 3' to the B1 element did not extend past the B1 element. Preliminarily, these results suggest that the B1 element is the 5' end of tumor associated kokopelli and that the 3'end needed to be isolated. Although, incomplete reverse transcriptase extension due to secondary structure caused by the B1 element could also explain the apparent 5' end truncation. Isolation of the normal 6.6knt transcript, at least the 3' end, would discern among these two possibilities. Experiments are ongoing to isolate kokopelli from normal kidney and liver cDNA libraries, part of task 2. Indeed, two clones have been isolated, one from each library, which give the expected pattern of hybridization on Northern blots. Sequencing of these clones is currently underway.

Isolation of chimeric transcripts using differential display analysis has not been vigorously exploited to date. Since a tumor specific clone with the predicted structure was isolated, I felt it necessary to fully analyze this clone (kokopelli) as it was expressed in every mouse mammary tumor tested as well as in a non-neoplastic immortal mammary cell line, experiments oulined in task 4. These results suggest that the alteration in expression pattern for kokopelli may be an early event in the progression of a normal cell to frank carcinoma. Differential display, with the modifications described above to increase specificity, may be a more direct method for the isolation of genuine chimeric transcripts. Therefore, more differential display experiments will be conducted over the next few months. Sequencing the amplicons isolated may yield a more direct answer as to chimerism than Northern analysis. Cloning PCR products using the T/A cloning vectors from Promega has been very successful and large amounts of sequence information can be generated quickly by Roswell Park's Biopolymer facility.

CONCLUSIONS

The isolation and characterization of oncogenes and tumor suppressor genes associated with breast cancer is a growing concern. Therefore, by using a mouse model system for the development of sporadic breast cancer I am in the process of isolating new cancer associated genes. To that end, Asch and Asch had previously found that endogenous retroviral-like elements were overexpressed in many mouse mammary tumors. These retroelements have the ability to reintegrate into the genome and cause mutation. Fortunately, the unique structure of the elements made it possible to use them as "molecular handles" to isolate new genes (for an example see 34). I have constructed a cDNA library from a DMBA-induced mouse mammary tumor which expressed elevated levels of IAPs. The library was screened with various probes specific to distinct regions of IAPs, the body of the element, the U3 region of the LTR and the U5 region of the LTR, to distinguish between IAP transcripts and those transcripts which were chimeric with the IAP LTR. The differential screen of the library proved successful insofar as I isolated clones which had the desired chimeric structure. However, I speculate that these transcripts arose due to recombination within the library, most likely during the amplification phase of library construction and not by retrotransposition. Serendipitously, one chimeric transcript did prove to be tumor associated. When used as a probe against total RNA from DMBA-induced tumors and normal tissues, a tumor specific transcript was detected. Northern analysis was extended to include tumors of other etiologies. The tumor associated transcript was observed in all mouse neoplasms examined. The tumor associated transcript was also observed in a non-neoplastic cell line, NOG-8 but not in any human neoplastic cell lines. Sequence analysis of the clone revealed a murine B1 element, a unique region, the U1 snRNP specific protein C, and env and LTR homologies. Extensive RT-PCR analysis has suggested that the structure obtained from the library up to but not including the env gene of IAP may exist within both tumor and normal cells. Attempts to isolate either the full length tumor associated transcript or the full length normal transcript by rescreening the library or using RACE techniques have proven difficult. However, two clones have been isolated recently which show promise. The clones were isolated from normal kidney and liver cDNA libraries and they hybridize to both normal and tumor associated transcripts (data not shown). These transcripts are currently being further characterized.

Another method of isolating the normal gene would be to determine genomic localization and then screen a chromosome specific library. The p17b clone (which I have renamed kokopelli) has been mapped using a Mus musculus X Mus spretus backcross obtained from the Jackson Labs. Backcross analysis has tentatively assigned kokopelli to chromosome 17 in the mouse, approximately 12.5 cM distal to the centromere. This result is useful in that many YAC contigs have been generated across this region and may be screened to isolate the genomic locus of kokopelli (Rosemarry Elliot, personal communication).

Differential display has been a powerful tool in visualizing the differences between two or more different cell types or tissues. I have attempted to isolate novel chimeric transcripts with this technique to circumvent potential problems associated with our library. The choice of an IAP specific anchored downstream primer has demonstrated differences between tumor tissue and normal mammary gland tissue from various physiological states. Isolation of tumor specific amplicons has led to the identification of potential IAP transcripts when these amplicons were used as probes on Northern blots. Other amplicons are being analyzed for their association with DMBA-induced mammary carcinogenesis.

The results to date suggest that isolation of chimeric clones using IAP sequences is possible. However, the technique has not worked exactly as anticipated. The isolated chimeric transcript may have arisen due to recombination within the library as is evident by the presence of two poly A tails within a single clone. Other methods of isolating chimeric transcripts is underway. Although the time schedule in the S.O.W is not being followed exactly, progress is being made. For example, a unique transcript has been isolated, sequenced, it's RNA expression pattern analyzed, it's genomic localization has been determined, and efforts are in progress to isolate the normal counterpart.

12

REFERENCES

1.) Kuff, E.L., and K.K. Lueders Ddv. Cancer Res. **51**, 183 (1988)

2.) Szyf, M., J. Theberge and V. Bozovic J. Biol. Chem. **270**, 12690 (1995)

3.) Goodchild, N.L.,D.A. Wilkinson and D.L. Mager Gene **121**, 287 (1992)

4.) Heidmann, O. and T. Heidmann Cell **64**, 159 (1991)

5.) Bera, T.K., D.K Panda, R.C. Guzman, L.-I. Hwand, T.P. Lin and S. Nandi Abst. 3368 in Proc. AACR **36**, 565 (1995)

6.) Ymer, S., W.Q.J. Tucker, H.D. Campbell and I.G.Young Nucl. Acids Res. **14**, 5901 (1986)

7.) Ymer, S., W.Q.J. Tucker, C.J. Sanderson, A.J.Hpael, H.D. Campbell and I.G.Young Nature **317**, 255 (1985)

8.) Cohen, J.B., T. Unger, G. Rechavi and D. Givol. Nature **306**, 797 (1983)

9.) Horowitz, M., S. Luria, G. Rechavi and D. Givol. EMBO J. **3**, 2937 (1984)

10.) Blatt, C., D. Aberdam, R. Schwartz and L. Sachs. EMBO J. **7**, 4283 (1988)

11.) Kittrell, F.S., C.J. Oborn and D. Medina. Cancer Res. **52**, 1924 (1992)

12.) Connelly, M.A., R.C.Grady, J.F. Mushinski and K.B. Marcu. Proc. Natl. Acad. Sci USA **91**, 1377 (1994)

13.) Kordon, E., G.H. Smith, R. Callahan and D. Gallahan.

14.) Michaud, E.J., M.J. van Vugt, S.J. Bultman, H.O. Sweet M.T. Davisson and R.P. Woychik Devel. **8**, 1463, (1994); Perry, W.L., N.G. Copeland and N.A. Jenkins. BioEssays **16**, 705 (1994); Chang-Yeh, A., D.E. Mold and R.C. Huang. Nucl. Acids Res. **19**, 3667 (1991); Van Houten, J., F. Natoli and B.B. Asch ; Banville, D., Y. Boie. Mol. Cell. Biol. **207**, 481 (1989)

15.) Asch, B.B. and H.L. Asch. Cancer Res. **50**, 2404 (1990)

16.) Asch, H.L., R.Z. Vinci, J.Y.Chang, F. Natoli, D.L. Stoler, G.R. Anderson and B.B. Asch. Cell Biol. Int. **17**, 961 (1993)

17.) Asch, B.B., H.L. Asch, D.L. Stoler and G.R. Anderson. Int. J. Cancer **54**, 813 (1993)

18.) Liang, P. and A.B. Pardee, Science 257, 967 (1992); Liang, P., L. Averboukh and A.B. Pardee, Nucl. Acids Res. **21**, 3269 (1993)

19.) Chisolm, D., BioTechnques **7**, 21 (1989)

20.) Troutt, A.B., M.G. McHeyzer-Williams, B. Pulendran and G.J.V. Nossal, Proc. Natl. Acad. Sci. USA **89**, 9823 (1992)

21.) Feuchter, A.E., J.D. Freeman and D.L. Mager. Genomics **13**, 1237 (1992); Feuchter-Murthy, A.E., J.D. Freeman and D.L. Mager. Nuc. Acids Res. **21**, 135 (1993)

22.) Yelick, P.C., C. Kozak, Y.K. Kwon, M.F. Seldin and N.B. Hecht, Genomics **11**, 687 (1991); Monson, J.M., J. Friedman and B.J. McCarthy, Mol. Cell. Biol. **2**, 1362 (1982); Bonnerot, C., N. Varin, C. Rabourdin-Combe, W.H. Fridman and J. Even, Nuc. Acids Res. **16**, 7180 (1988); Oberbaumer, I. and U. Magdolen, *ibid*, 7181 (1988)

Figure 1. cDNA Library Screen. The cartoon is taken from Feuchter et.al. (34) and depicts the possible chimeric transcripts that can be isolated. A). The endogenous IAP element. The thick black lines represent the flanking cellular DNA. The LTRs are shown as boxes with the U3,R, and U5 regions designated. The thin line represents the body of the element. Probes to the body of the element, the U3 region and the U5 region (the stippled box, the hatched box and the crossed box, respectively) are described in Methods and Materials. B.-D). Representation of the possible genomic structures with corresponding mRNA transcripts (dashed lines) and the probe(s) which would recognize the transcripts.

Figure 2. Northern Analysis of p13a. Northern analysis was done as describe in Methods and Materials. A). Ten micrograms of RNA from the DMBA-induced tumors was hybridized with the p13a clone isolated from the library. The transcripts which run at 7.2knt and 5.4knt are cross hybridization with the IAP sequences within the probe. The 3.5knt transcript represents the unique region of p13a. B). rDNA was used as a loading control.

Figure 3. Northern Analysis of p17b (kokopelli). A). Schematic representation of the p17b clone isolated from the tumor specific cDNA library. The open box represents the unique region while the checkered box represents IAP sequences. The box with circles is the 5' end unique region, the hatched box represents the murine B1 repetitive element, the stippled box denotes the other unique region and the box with the diamonds is the U1 snRNP Specific protein C. B). Ten micrograms of RNA from DMBA-induced tumors generated in our lab (T6.8-T12.21) and from Dan Medina (4657-6821) along with various normal tissues were probe with the 1.3kb probe which represents the non-IAP portion of the clone. The 6.6knt transcript is expressed in all tissues while the 1.4knt transcript is expressed only in the tumors. C). rDNA was used as a loading control.

Figure 4. Sequence and Northern analysis of p17b-20 (U1 snRNP Specific Protein C). A). Sequence analysis of clone p17b-20 (lower strand) shows complete homology to the splicing factor U1 snRNP Specific Protein C (upper strand) as determined by searching the GCG data base. The sequence of p17b-20 was generated by using the M13 universal reverse primer and was not sequenced beyond what is shown. B). Northern analysis using p17b-20 as a probe on several DMBA-induced mammary carcinomas, many normal mouse tissues, cell lines as well as two human tumor cell line (MDA-231 and MCF-7). There was no altered expression pattern of p17b-20 as compared to a ribosomal gene probe, L18 (C).

Figure 5. Analysis of clone p48. A). Sequence analysis revealed a chimeric molecule between two different cellular transcripts. The clone demonstrated two apparent poly A tails, positions 54 to 71 and 305 to 330. However, upon closer examination the first poly A tail is not preceded by a poly A signal (AAUAAA). Indeed a data base search on the sequence between the two stretches of A's revealed high homology to positions 1672 to 1909 of a very large (16303bp) hydrophobic mitochondrial protein gene (position 54 to 292). This type of chimera was detected previously except with IAP sequences and a unique cellular transcript (compare with Fig 6A). This sequence is the non-coding strand and was generated in Roswell Park's Biopolymer core facility using the ABI automated DNA sequencer. B). Northern analysis detects an altered transcript in a few DMBA-induced tumors. The probe recognized a 1.3knt transcript in normal tissues and in most tumors but detects a transcript of 1.0knt only in tumors 11.10A, 12.21 and 5978. C). Sequence homology searches reveals complete homology to smooth muscle alpha actin. Smooth muscle alpha actin is the top strand and clone p48 is the bottom strand.

Figure 6. Sequence analysis of 3'end of clone p17b. The 1.8kb clone representing p17b was sequenced and it was found to be chimeric with IAP. Note the two polyadenylation signals (AATAAA) (underlined) and two poly A tails, one associated with the snRNP and the other associated with the IAP LTR. The nucleotide position where p17b homology to the IAP is identicle is marked in outline. The arrow indicates the boundry between the IAP env gene and the LTR. For clarity, only the 3' 350 bp of p17b is shown.

Figure 7. Sequence comparison of different clones isolated by 5' and 3' RACE. A). 5' RACE clones isolated by PCR with vector primers and specific p17b primers were ligated into pGEM T/A Easy and sequenced. p17bu322 is the sequence of the p17b clone isolated by library screening. The sequence starts at position 280. There PCR amplicons representing the 5' end of the clone are shown below p17b (Kpli-500, Kpli3 and Kpli2r). Kpli is the consensus sequence of all four clones. There is considerable homology between all four independently isolated clones when minor sequencing errors are taken into account. B). The same technique was used to isolate the 3' end of p17b as well. Sequence homology between the two clones is identical, again taking into account minor sequencing errors. Interestingly, the sequence shown extends from within the unique region and through snRNP specific protein C.

Figure 8. Northern analysis of p17b (kokopelli). A). Different mouse mammary tumors were run on a Northern gel and probed for the presence of the 1.4knt tumor specific transcript found in the DMBA-induced tumors. NOG-8 is a transformed, non-neoplastic cell line, the 1.4knt transcript is detectable but at very low levels. fC3H is a mammary tumor from BALB/cJ mice foster nursed on C3H mice to infect the BALB/cJ with MMTV. After several rounds of pregnancy the animals develop mammary tumors. D1, D2, Dim3 and C4 are all hyperplastic alveolar nodules (HANs) which have been serially transplanted in isogenic mice. These HANs eventually go on to develop tumors. Comma-D is a mouse mammary carcinoma cell line. MDA-231, T47-D, MCF-7 and SKBR-3 are all human mammary carcinoma cell lines. Kokopelli has not yet hybridized to any human cell line or tissue. B). Tissue distribution of the larger 6.6knt transcript, the fC3H tumor serves as the positive control. The 6.6knt transcript was detected in all mouse tissues examined, albeit to varying levels. Upon longer exposure of the gel, transcripts are indeed detected in whole blood, liver and stomach. The 1.4knt transcript appears to be absent in all non-neoplastic tissues. The ribosomal 28S and 18S bands are marked in each figure. Ethidium staining of the gel reveals approximately equal loads (not shown).
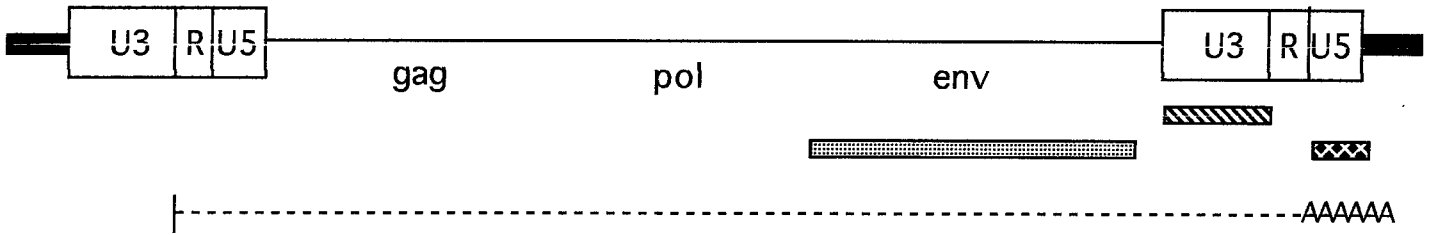
Figure 9. Chromosomal localization of kokopelli. Analysis of Mus musculus X Mus spretus backcross localizes kokopelli to chromosome 17, approximately 12.5 cM distal from the centromere. The other markers on chromosome 17 have been omitted for clarity.

Figure 10. Differential display. An IAP specific primer was used in the reverse transcription reaction to amplify IAP specific gene sequences. A). Representation of differential display using the IAPU3S2 downstream primer in conjunction with the AP-3, AP-4 and AP-5 primers from GeneHunter. The arrowheads indicates some of the bands which were isolated and used as a probes for Northern analysis. B). Same as in A except that AP-1m through AP-5m were used as upstream primers (see Methods and Materials for details on primers). The bands which were isolated have not yet been used as probes against Northern strip blots at the time of writing this report.
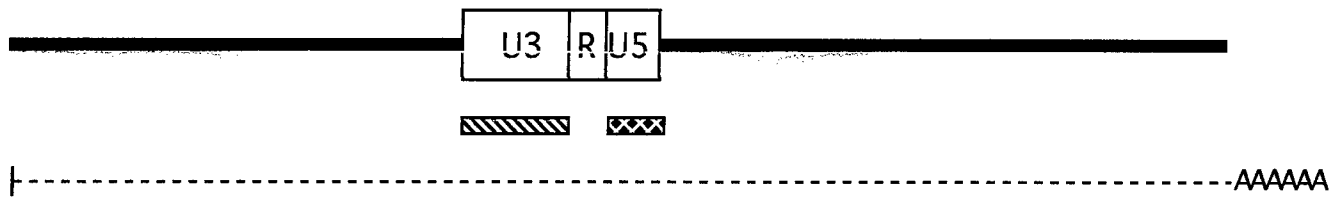
Figure 11. Northern analysis of DD products. Amplicons from AP-3/IAPU3S2 were labeled with a-32P and used as probes on Northern strip blots containing RNA form liver and T6.8. A). Amplicon 2. B). Amplicon 3. C). Amplicon 4. D). Amplicon 5. Note that amplicons 2 and 3 hybridize to bands of 7.2knt and 5.4knt. This hybridization pattern is seen with the full length IAP molecule as a probe. Therefore these clones were considered to be IAP related sequences. The ribosomal RNA bands (28S and 18S) are marked. Hybridization of amplicon 4 was not different between liver and tumor (arrows in C)
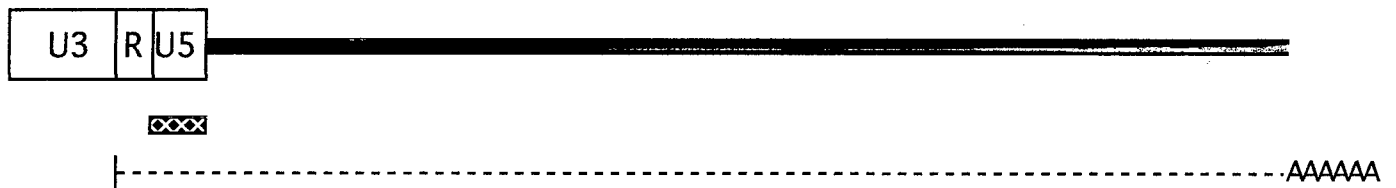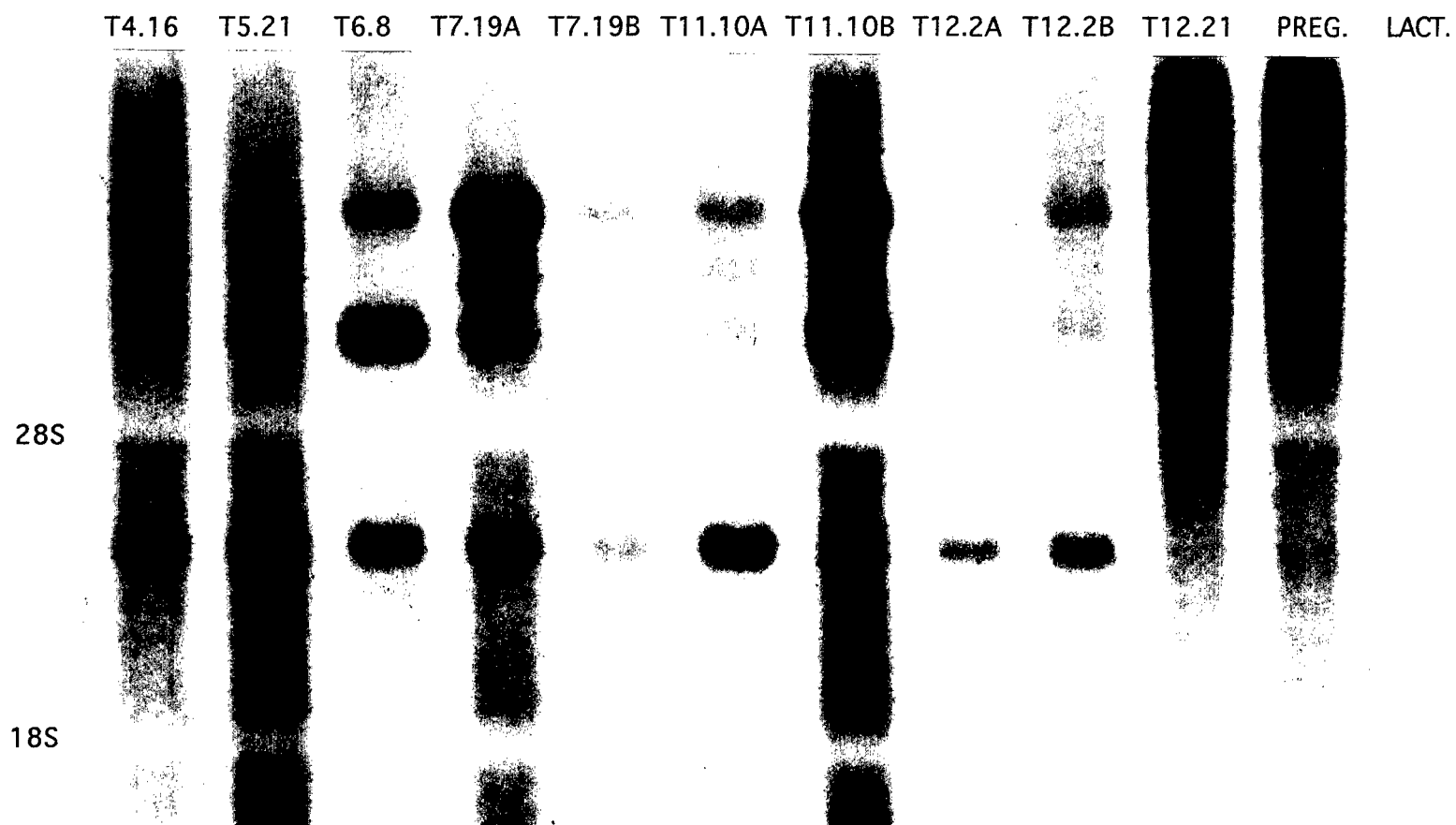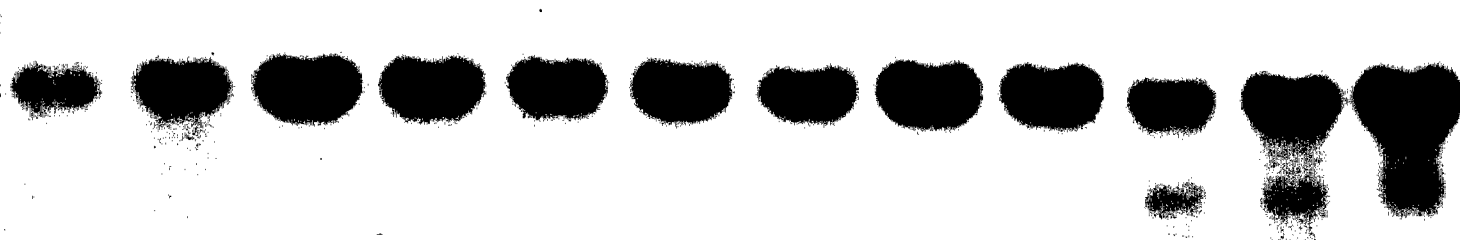
Figure 1

Figure 2

A.

| T4.16 | T5.21 | T6.8 | T7.19A | T7.19B | T11.10A | T11.10B | T12.2A | T12.2B | T12.21 | PREG. | LACT. |

28S

18S

B.

Figure 3

A.

B.

28S

18S

C.
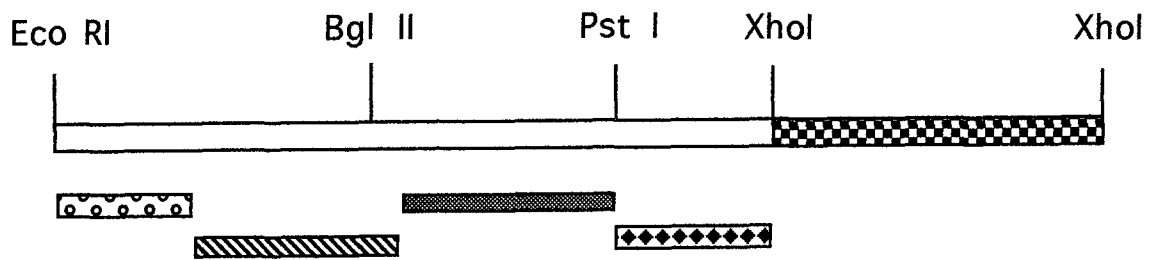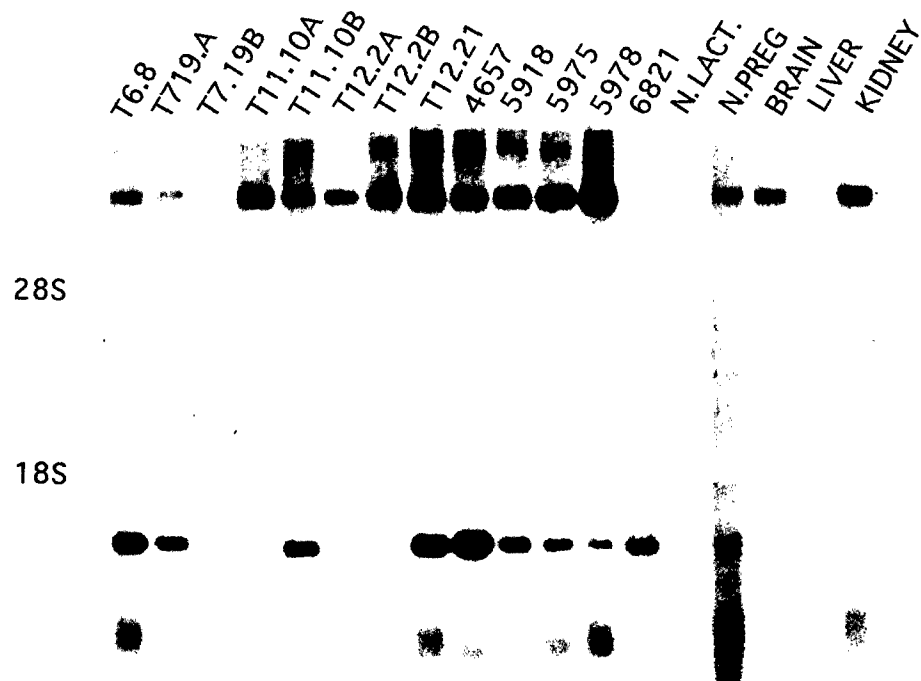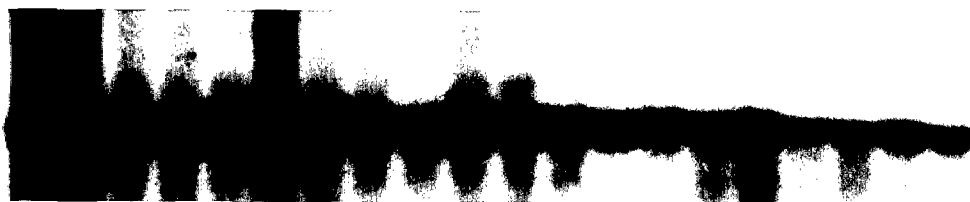
Figure 4

A.

Mmsnrnp GGATCGGGGT AGCCAACGGT TGTGGAGCAT CATGCCCAAG TTTTATTGTG

Mmsnrnp ACTACTGTGA TACGTATCTT ACCCACGATT CTCCATCTGT GAGGAAGACA
P17b-20   CGATNTGT GAGGAAGACA

Mmsnrnp CACTGCAGTG GTCGGAAACA CAAAGAGAAT GTGAAAGACT ATTATCAGAA
P17b-20 CACTGCAGTG GTCGGAAACA CAAAGAGAAT GTGAAAGANT ATTATCAGAA

Mmsnrnp ATGGATGGAA GAGCAGGCCC AGAGCCTGAT TGACAAGACA ACGGCTGCAT
P17b-20 ATGGATGGAA GAGCAGGCCC AGAGCNTGAT TGACAAGACA ACGGNTGCAT

Mmsnrnp TTCAACAAGG GAAGATCCCT CCTGCTCCGT TCTCTGCTCC TCCGCCTGCA
P17b-20 TTCAACAAGG GAAGATCCTT CCTGTTCCGT TTTNTGNTCC TCCGCCTGCA

Mmsnrnp GGGGCCATGA TCCCACCTCC CCCCAGTCTC CCGGGCCCTC CTCGGCCTGG
P17b-20 GGGGCCATGA TCCCACTTCC CCCCAGTTTC CCGGGCCTTC TTCGGCNTGG

Mmsnrnp CATGATGCCT GCCCCCCACA TGGGAGGCCC TCCCATGATG CCAATGATGG
P17b-20 CATGATGCCT GCCCCCCACA TGGGAGGCCC TCCCATGATG CCAATGATGG

Mmsnrnp GCCCCCCTCC GCCCGGAATG ATGCCCGTGG GACCAGCTCC TGGGATGAGA
P17b-20 GCCCCCCTCC GCCCGGAATG ATGCCCGTGG GACCAGCTCC TGGGATGAGA

Mmsnrnp CCACCCATGG GAGGCCACAT GCCCATGATG CCCGGACCTC CCATGATGAG
P17b-20 CCACCCATGG GAGGCCACAT GCCCATGATG CCCGGACCTC CCATGATGAG

Mmsnrnp ACCTCCTGCC CGCCCTATGA TGGTGCCCAC CCGGCCTGGC ATGACCCGGC
P17b-20 ACCTCCTGCC CGCCCTATGA TGGTGCCCAC CCGGCCTGGC ATGACCCGGC

Mmsnrnp CAGACAGATA AGAGCAGAAG CGCTCTTGAT GGTTTTGTAT TTCTTGTTCT
P17b-20 CAGACAGATA AGAGCAGAAG CGCTCTTGAT GGTTTTGTAT TTCTTGTTCT

Mmsnrnp GTTCCACCAG GAGCTCTTGG TGCTGAGCCC GAGTGTTTAC TAGATGCATG
P17b-20 GTTCCACCAG GAGCTCTTGG TGCTGAGCCC GA.TGTTTAC TAGATGCATG

Mmsnrnp GAAAGGAAAC TTCCCTTCCT AACTGAATAT TTTTGGAGGG AGAAATAATA
P17b-20 GAAAGGAAAC TTCCCTTCCT AACTGAATAT TTTTGGAGGG AGAAATAATA

Mmsnrnp CAAAAAAGTG CAGTTTTCAC TTATATTGTG AAATGTGAAA ATAAAGTCAT
P17b-20 CAAAAA.GTG CAGTTTTCAC TTATATTGTG AAATGTGAAA ATAAAGTCAT

Mmsnrnp CAGCTCTTTT AGTTAAAAAAAAA
P17b-20 CAGCTCTTT . AGTTAAAAAAAA

B.

C.

MCF-7 CELLS
MDA-231 CELLS
MDA-231 CELLS
COMMA D CELLS
NOG-8 CELLS
DMBA CELLS
UTERUS
HEART
SPLEEN
KIDNEY
PREGNANT
LACTATING
6821
VIRGIN
5978
4657
T12.2A
T11.10A
T7.19A
T6.8

20

Figure 5

A.

ANGANTGTATACGACTCACTATAGGGCGAATGGGTACCGGGCCCCCCTCGAGTTTTTTTTTTTTTTTTGGATCTTTCCTTTAGGCATTCCGGTGTTGGGTTAACAGAGAAGTTATAGGTGG/
10   20   30   40   50   60   70   80   90   100   110   120

ATTTATTTATAGATTAGTGTGATTATTGCCTATAGTCTGATTAACTAACAATGGTTATCCGAGTTGTTATACGCGTTGTTATGCCTGGAGAATTGGAATTCTTGTTACTCATACTAACAGTGTTGCATCT
130   140   150   160   170   180   190   200   210   220   230   240

TATAAAGTTATAGATTAACCCAATTTAAGTTTAGGAAGTTGGTGTAAATCTCGTGTCGAGTTTTTTTTTTTTTTTAAGGGTAGCACATGTCTGATTTATTTTTAAA
250   260   270   280   290   300   310   320   330   340   350   360

AAAACCACGANTAACAAATCAAAGCTTTGGGCAGGAATGATTTGGAAAGGAACTGGAGGGCCTGATCCACAAAAACGTTCACAGTTGTNTGCTAAANGCANANCAGGGGGGACTTAN
370   380   390   400   410   420   430   440   450   460   470

21

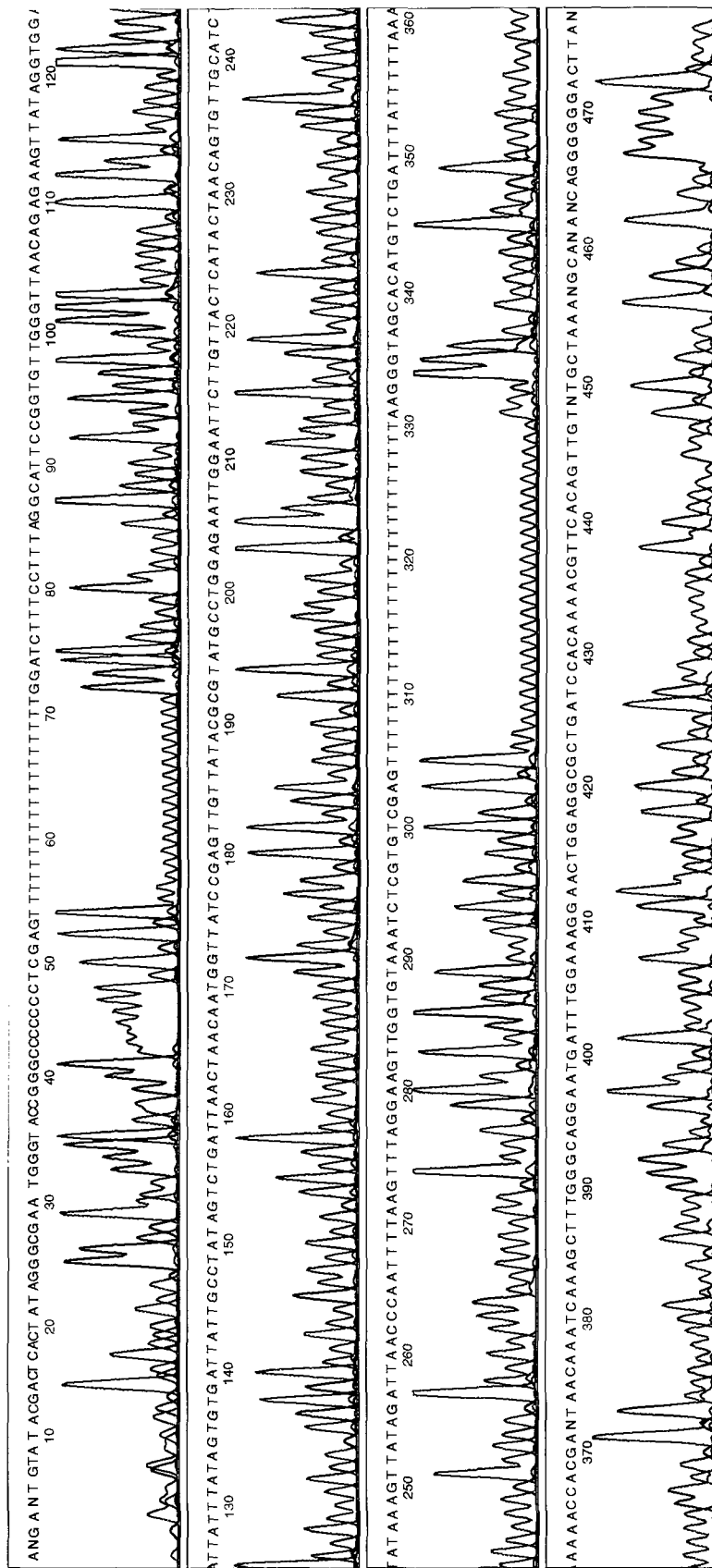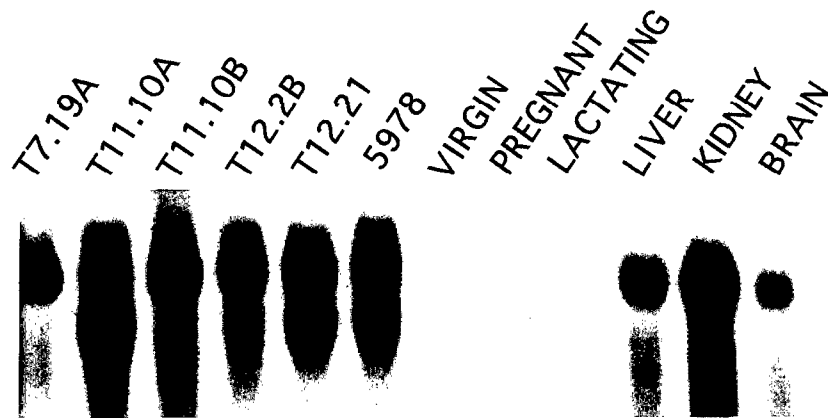# Figure 5

B.

C.

```
              801
α-Actin  CCATTGGAAA CGAACGCTTC CGCTGCCCAG AGACTCTCTT CCAGCCATCT


α-Actin  TTCATTGGGA TGGAGTCAGC GGGCATCCAC GAAACCACCT ATAACAGCAT
P48                                    AAACCACCT ATAACGGCAT


α-Actin  CATGAAGTGT GATATTGACA TCAGGAAGGA TCTCTATGCT AACAACGTCC
P48      CATGAAGTGT GATATNGACN TCAGGAAGGA TTTTTATGNT AACAACGTCC


α-Actin  TGTCAGGGGG TACCACCATG TACCCAGGCA TTGCTGACAG GATGCAGAAG
P48      TGTCAGGGGG TACCACCATG TACCCAGGCA TTGNTGACAG GATGCAGAAG


α-Actin  GAGATCACAG CCCTCGCACC CAGCACCATG AAGATCAAGA TCATTGCCCC
P48      GAGATAACAG CCTTNGCACC CAGCACCATG AAGATCAAGA TCATTGCCCT


α-Actin  TCCAGAACGC AAGTACTCTG TCTGGATCGG TGGCTCCATC TTGGCTTCGC
P48      TCCAGAACGC AAGTANTCTG TNTGGATCGG TGGCTCCATC CTGGCTTCGC


α-Actin  TGTCTACCTT CCAGCAGATG TGGATCAGCA AACAGGAATA CGACGAAGCT
P48      TGTTTACCTT CCAGCAGATG TGGATCAGCA AACAGGAATA CGACGAAGNT


α-Actin  GGGCCCTCCA TCGTCCACCG CAAATGCTTC TAAGTCCCCC CTGCTCTGCC
P48      GGGCCCTCCA TNGTCCACCG CAAATGNTTN TAAGTCCCCC CTGNTNTGCN


α-Actin  TCTAGCACAC AACTGTGAAC GTTTTGTGGA TCAGCGCCTC CAGTTCCTTT
P48      TNTAGCANAC AACTGTGAAG CTTTTGTGGA TCAGCGCCTC CAGTTCCTTT


α-Actin  CCAAATCATT CCTGCCCAAA GCTTTGATTT GTTACTCGTG GTTTTTTTTA
P48      CCAAATCATT CCTGCCCAAA GCTTTGATTT GTTANTCGTG GTTTTTTTNA


α-Actin  AAAATAAATC AGACATGTGC TACCCTTAAC TTGTAAAAAA
P48      AAAATAAATC AGACATGTGC TACCCTT.................AAAA AAAAAAAAAA
```

Figure 6

P17B AGGGAGAAAT AATACAAAAA AGTGCAGTTT TCACTTATAT TGTGAAATGT
IAP   TGGCTCCTTT GACCTACACA CTGGGGATTT GACCTCTATC TCCACTCTCA

P17B TGAAAA.... TAAAGTCATC AGCTCTTTTA GTTAAAAAAA AAAAAAAAAA
IAP   TTAATATGGG TGGGCCTATT TGCTCTTATT AAAAGGAAAG GGGGAGATGT
                                                        ⇑

P17B CTCGAGACTA GCCCACATTC GCCGTTACAA GATGGCG.TG ACAGCTGTGT
IAP   TGGGAGCCGC GCCCACATTC GCCGTTACAA GATGGCGCTG ACAGCTGTGT

P17B TCTAAGTGGT AAACAAATAA TCTGCGCATA TGCCGAGGGT GGTTCTCTAC
IAP   TCTAAGTGGT AAACAAATAA TCTGCGCATA TGCCGAGGGT GGTTCTCTAC

P17B TCCATGTGCT CTGCCTTCCC CGTGACGTCA ACTCGGCCGA TGGGCTGCAG
IAP   TCCATGTGCT CTGCCTTCCC CGTGACGTCA ACTCGGCCGA TGGGCTGCAG

P17B CCAATCAGGG AGTGACACGT CCTAGGCGAA ATATAACTCT CCTAAAAAAG
IAP   CCAATCAGGG AGTGACACGT CCTAGGCGAA ATATAACTCT CCTAAAAAAG

P17B GGACGGGGTT TCGTTTTCTC TCTCTCTTGC TTCTTACACT CTTGCTCCTG
IAP   GGACGGGGTT TCGTTTTCTC TCTCTCTTGC TTCTTACACT CTTGCTCCTG

P17B AAGATGTAAG CAATAAAG.T TTGCCGCAGA AGAAAAAAAA AAAAAAAAAA
IAP   AAGATGTAAG CAATAAAGTT TTGCCGCAGA AGATTCTGGT CTGTGGTGTT

Figure 7

A

```
P17bu322  TCATTTCCNN GTAATTCCCA GTATGANTCC CAGCAGGCAG GCATGGGGAG
Kpli-500                        ATGACTCC CAGCAGGCAG GCATGGG.AG

Kpli2r            CGA GGAATTCCGA TGACNCCCCA GCAGGCC.AG GCATGGG.AT
Kpli           TCATTTCCga G.AATTCC.A ..AtgactCc cagcaggcAG GCATGGG.Ag


P17bu322  TGCACGCCTT AAATTTCAGC ANTCTGGGGC AGAAGTAGGT GGATNTCTGT
Kpli-500  TGCACGCCTT AAATCTCAGC ACTCTGGGGC AGAAGTAGGT GGATCTCTGT
Kpli3                                     GTAGGT GGATCTCTGT
Kpli2r    GTCACGCCTT AAATCTCAGC ACTCTGGGGC AGAAGTAGGT GGATCTCTGT
Kpli      tgCACGCCTT AAATcTCAGC Ac TCTGGGGC AGAAGTAGGT GGATcTCTGT


P17bu322  GTTGAGGCCA GCNTGGTCTA CAGAACCAGT TACAGGATAG CCAGGGCTTT
Kpli-500   GTTGAG .CCA GCCTGGTCTA CAGAACCAGT TACAGGATAG CCAGGGCTTT
Kpli3     GTTGAG. CCA GCCTGGTCTA CAGAACCAGT TACAGGATAG CCAGGGNTTT
Kpli2r    GTTGAGGCCA GCCTG. TCTA CAGAACCAGT TACAGGATAG C. A
Kpli      GTTGAG.CCA GCcTGgTcTA CAGAACCAGT TACAGGATAG CcAGGGcTTT


P17bu322  ATAATAGAGA GATTTATTCC AAACAAAAAC AACAGCAATA ACAGAAAACA
Kpli-500   ATAATAGAGA GATCTATTCC AAACAAAA. C AACAGCAATA ACAGAAAACA
Kpli3     ATAATAGAGA GATCTATTCC AAACAAAAAC AACAGCAATA ACAGAAAACA

Kpli      ATAATAGAGA GATcTATTCC AAACAAAAaC AACAGCAATA ACAGAAAACA


P17bu322  AAACAAAACA AAACAACAAA ACCCACAGCC AGAACTAACC AAACAGAAAC
Kpli-500   AAACAAAACA AAACAACAAA ACCCACAGCC AGAACTAACC AAACAGAAAA
Kpli3     AAACAAAACA AAACAACAAA ACCCACAGCC AGAACTAACC AAACAGAAAC

Kpli      AAACAAAACA AAACAACAAA ACCCACAGCC AGAACTAACC AAACAGAAAc


P17bu322 TGAGCAAGCA GAGCTGGTGT GCCTGTTGCC GTGTGAGACA A

Kpli3     TGAGCAAGCA GAGCTGGTGT GCCTGTTGCC GTGTGAG

Kpli      TGAGCAAGCA GAGCTGGTGT GCCTGTTGCC GTGTGAGACA A
```

24

Figure 7

B

```
Kpli3-3   TAGTGATTCA TGTGTGTTGT CACCGTCCTC .CCCAGGCCC CCCAAAAAAA
P17bu322  TAGTTTGTCA TGTGTGTTGT CACCGTCCTC CCCCATTCCC CCCAAAAAAA


Kpli3-3   CCCAAACAAA CAGAGTTTCC CTGTGTATAG CCCTGGCTGT CCTGGAACTT
P17bu322  CCCAAACAAA CAGAGTTTCC CTGTGTATAG CCCTGGCTGT CCTGGAACTT


Kpli3-3   ACTCTGTAGA CAAGGCTGGC CCAGAATGCA AAGATTTTGT GTTGAGTATT
P17bu322  ACTCTGTAGA CAAGGCTGGC CCAGAATGCA AAGATTTTGT GTTGAGTATT


Kpli3-3   GGGACTAAAG GACCCCCCCC CCTTTTTAGG TTTAATGGAC GTCCCGTCTG
P17bu322  GGGATTAAAG GACCCCCCCC CCTTTTTAGG TTTAATGGAC GTCCCGTCTG


Kpli3-3   CTGTC..ACCT GAATGTTGTT TCTCACTCCT TTCCCTTTGT TCTGTTCTGC
P17bu322  CTGTCAACCT GAATGTTGTT TCTCACTCCT TTCCCTTTGT TCTGTTCTGC


Kpli3-3   AGCTCCTGGG AGT .GACCAC CCATGGGAGG CCACATGCCC ATGATGCCCG
P17bu322  AGCTCCTGGG ATGAGACCAC CCATGGGAGG CCACATGCCC ATGATGCCCG


Kpli3-3   GACCTCCCAT GATGAGACCT CCTGCCCGCC CTATGATGGT GCCCACCCGG
P17bu322  GACCTCCCAT GATGAGACCT CCTGCCCGCC CTATGATGGT GCCCACCCGG


Kpli3-3   CCTGGCATGA CCCGGCCAGA CAGATAAGAG CAGAAGCGCT CTTGATGGTT
P17bu322  CCTGGCATGA CCCGGCCAGA CAGATAAGAG CAGTTGCGCT CTTGATGGTT


Kpli3-3   TTGTATTTCT TGTTCTGTTC CACCAGGAGC TCTTGGTGCT GAGCCCGAGT
P17bu322  TTGTATTTCT TGTTCTGTTC CACCAGGAGC TCTTGGTGCT GAGCCCGAGT


Kpli3-3   GTTTACTAGA TGCATGGAAA GGAAACTTCC C
P17bu322  GTTTACTAGA TGCATGGAAA GGAAACTTCC C
```
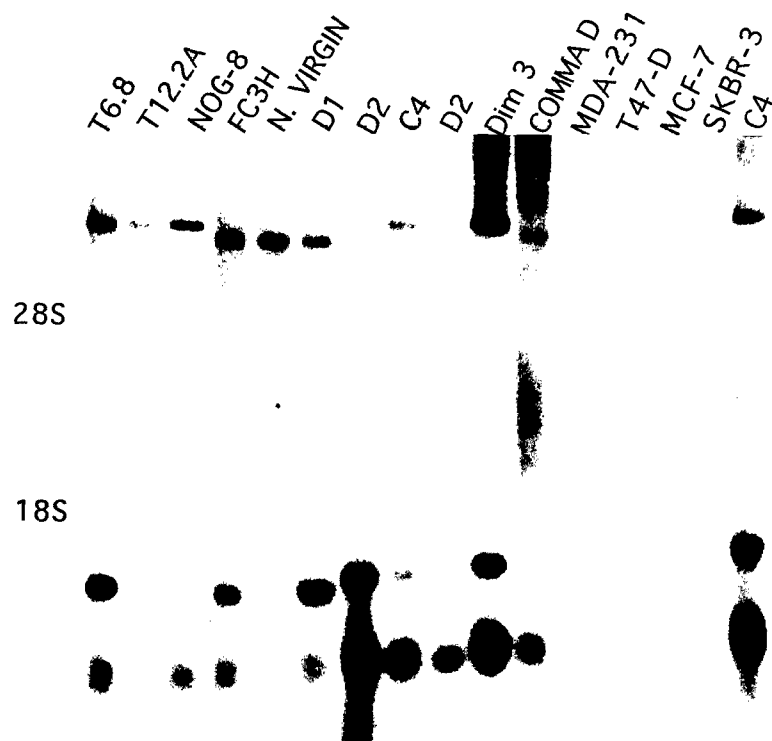
25

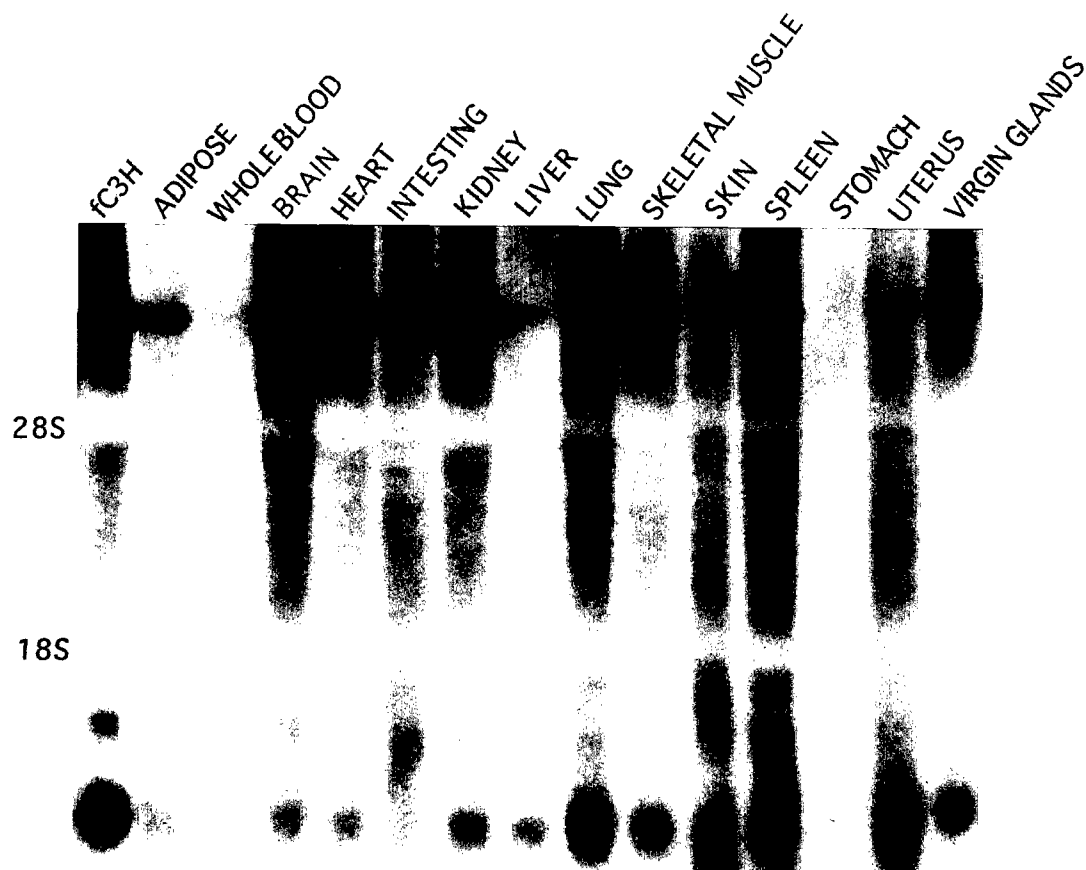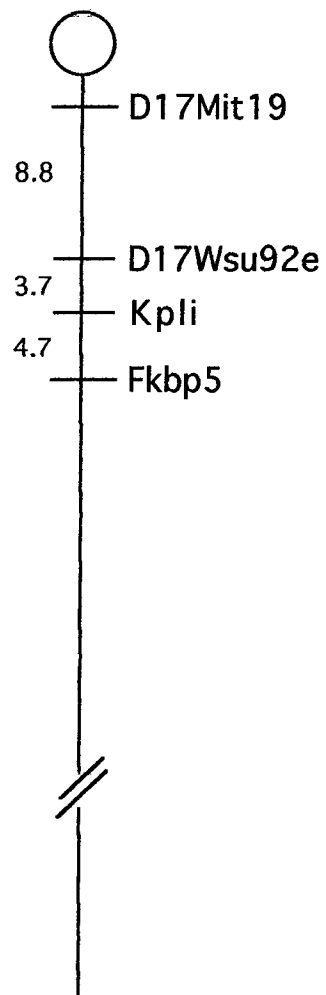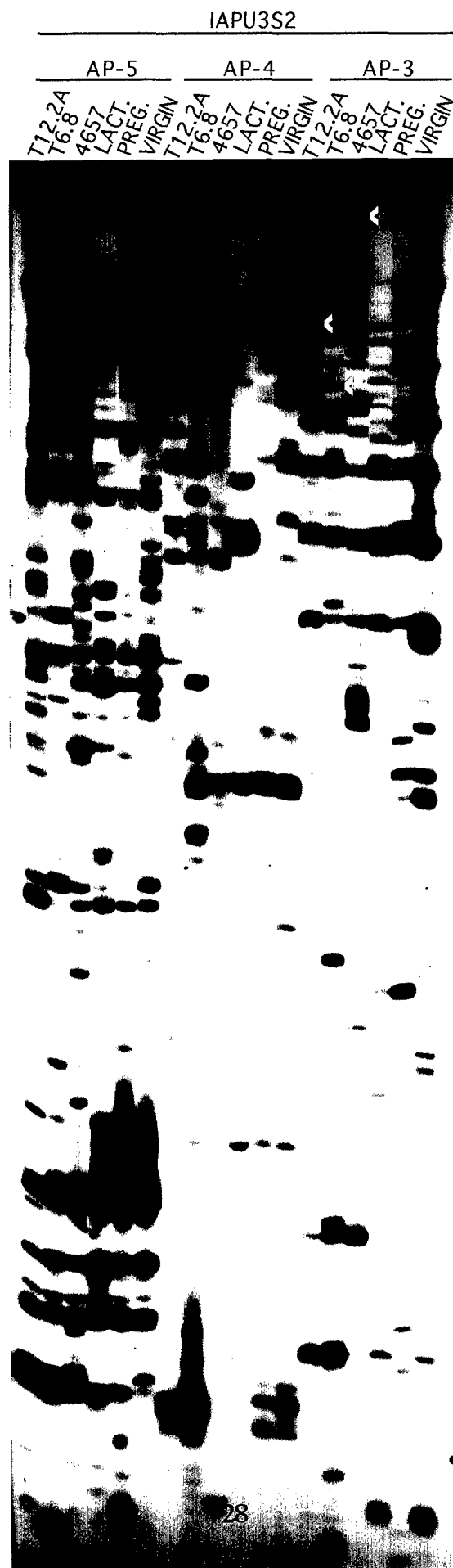# Figure 8

A.

28S

18S

B.

28S

18S

Figure 9

D17Mit19

8.8

D17Wsu92e

3.7

Kpli

4.7

Fkbp5

27

Figure 10

A.

Figure 10

B.

Figure 11